

## CLAIMS

What is claimed is:

1. A system for tuning the text-to-speech conversion process, the system comprising:

a text-to-speech engine, said text-to-speech engine receiving at least one text-input and converting said text-input into a processed representation,

said processed representation including at least one speech feature associated with at least one segment of said representation; and

a visual editing interface, said visual editing interface displaying said processed representation using at least one graphical indicator on an output device, wherein said segment is displayed on said output device using said graphical indicator corresponding to said speech feature.

2. The system of Claim 1 wherein said visual editing interface provides at least one editing function to a user, the editing function enabling the modification of said speech feature associated with said segment through a change in the corresponding said graphical indicator.

3. The system of Claim 2 wherein said visual editing interface associates said speech feature corresponding to said segment with said graphical indicator, wherein the user's modification of said graphical indicator results in a corresponding change in said speech feature of said segment.

4. The system of Claim 1 wherein said speech feature is at least one of the following: normalized text, part-of-speech, parsing of text, chunking of text, boundary strength, pause duration, transcription, speech rate, syllable duration, segment duration, pitch, word prominence, emphasis, formant mixing mode, unit selection override, intensity contour, formant trajectories, and allophone rules.

5. The system of Claim 1 wherein said graphical indicator comprises at least one of the following: graphical style, font faces, coloring, vertical spacing, horizontal spacing, italicization, boldness, underlining, blinking, crossing-out, text orientation, text rotation, punctuation symbols and graphical symbols.

6. The system of Claim 1 wherein said processed representation employs a parameterized aligned sound records format.

7. The system of Claim 1 wherein said segment comprises at least one of the following: word, letter, syllable, pause, word boundary and punctuation-mark.

8. The system of Claim 1 wherein said visual editing interface operates as a plug-in for a graphical user interface.

9. The system of Claim 8 wherein said plug-in is an ActiveX control.
10. The system of Claim 1 wherein said visual editing interface allows editing of said input-text wherein said input-text contains at least one non-editable said text segment and at least one editable said segment.
11. The system of Claim 1 wherein said visual editing interface is language independent.
12. The system of Claim 1 wherein said visual editing interface provides the user with speech audio output of said processed representation.
13. The system of Claim 1 wherein visual editing interface is connected to a data-store for storing and retrieving said representation.
14. The system of Claim 1 wherein the said processed representation is a textual representation.
15. The system of Claim 14 wherein the said textual representation is used to generate said processed representation.
16. The system of Claim 15 wherein said textual representation is stored and accessed from a data store.

17. The system of Claim 14 wherein said textual representation is used to generate synthesized speech using a TTS system distinct from said text-to-speech engine.

18. A system for providing a text-to-speech interface, the system comprising:

a visual interface connected to a text-to-speech engine; and  
at least one communication channel connecting said visual interface to said text-to-speech engine, said text-to-speech engine communicating with said visual interface over said communication channel by sending and receiving at least one data segment in a format.

19. The system of claim 18 wherein said format of said data segment is a parameterized aligned sound records format.

20. The system of claim 18 wherein said text-to-speech engine sends said data segment in the parameterized aligned sound records format to said visual interface, said visual interface rendering said data segment in a visual form, said visual interface allowing editing of said data segment to produce an edited data segment, said visual interface sending said edited data segment to said text-to-speech engine.

21. The system of claim 18 wherein said visual interface sends data to said text-to-speech engine over a first said communication channel and said text-to-speech engine sends data to said visual interface over a second said communication channel.

22. A method for visual tuning text-to-speech conversion process, the method comprising:

converting an input-text to a processed representation using a text-to-speech engine, said processed representation including at least one speech feature of said input-text;

displaying said processed representation on a visual editing interface connected to said text-to-speech engine, said speech feature of said processed representation being displayed in a corresponding graphical form; and

providing an editing function in said visual editing interface to a user for modifying said speech feature in said graphical form.

23. The method of Claim 22 further comprising:

generating speech audio equivalent of said processed representation through said visual editing interface.

24. The method of Claim 22 further comprising:

saving said processed representation in a data store; and

loading said processed representation stored in said data store into said visual editing interface.

25. The method of Claim 22 further comprising:  
converting said processed representation into a textual representation.

26. The method of Claim 25 further comprising:  
converting said textual representation into a processed representation.

27. The method of Claim 25 further comprising:  
storing said textual representation in a data store; and  
loading said textual representation stored in said data store into said visual editing interface.

28. The method of Claim 25 further comprising:  
using said textual representation to synthesize speech using a TTS system distinct from said text-to-speech engine.